

DOCUMENT RESUME

ED 265 516

CS 008 291

AUTHOR Power, Michael A.; White, Thomas G.
TITLE Measures of Word Frequency as Predictors of English Root Word Knowledge.
PUB DATE Dec 85
NOTE 17p.; Paper presented at the Annual Meeting of the National Reading Conference (35th, San Diego, CA, December 3-7, 1985).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Education; Grade 3; Grade 4; Morphology (Languages); *Predictive Validity; *Reading Research; *Vocabulary Development; *Vocabulary Skills; *Word Frequency; *Word Recognition
IDENTIFIERS *Root Words

ABSTRACT

To determine if measures of the frequency of English root words can serve as reliable predictors of knowledge of those words, a 98-item multiple choice vocabulary test, in which all of the items were English root words, was administered to 46 third graders and 45 fourth graders. The percentage of correct responses for the words was correlated with four measures of word frequency, using two methods of grouping--arithmetic and logarithmic. Three hypotheses were tested: (1) for morphologically basic (unaffixed) root words, there should be a strong relationship between word frequency and word knowledge; (2) the advantage of logarithmic grouping over arithmetic grouping should be less apparent, or non-existent, for morphologically basic words; and (3) for grouped basic words, simple item frequency would prove as strong a predictor of word knowledge as family frequency. All three of the hypotheses were confirmed. All frequency by knowledge correlations were quite high, arithmetic grouping gave the highest correlations, and simple item frequency proved to be the strongest predictor of root word knowledge.
(Author/HOD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

X This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

ED265516

Measures of Word Frequency as Predictors
of English Root Word Knowledge

Michael A. Power & Thomas G. White

Center for the Development of Early Education

Kamehameha Schools

Honolulu, Hawaii 96817

Paper presented to the National Reading Conference; San Diego,
California, December, 1985.

PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Michael A. Power

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC).

CS 008 291

Abstract

A 98 item multiple choice vocabulary test, in which all of the items were English root words, was administered to third and fourth grade children. The percentage of correct responses for the words was correlated with four measures of word frequency, using two methods of grouping, arithmetic and logarithmic.

Previous studies of the relationship between word frequency and children's word knowledge have shown low to negligible correlation. Graves, Ryder, Slater, & Calfee (1983) however, found relatively strong correlations, provided the words were grouped by frequency. In addition, they found a high correlation between word knowledge and "family frequency."

We hypothesize that much of the unpredictability in the frequency-knowledge relationship is due to variation in children's knowledge of prefixes and suffixes. On this basis we made three predictions. (1) For morphologically basic (unaffixed) root words, there should be a strong relationship between word frequency and word knowledge. (2) The advantage of logarithmic grouping over arithmetic grouping should be less apparent, or non-existent, for morphologically basic words. (3) For grouped basic words, simple item frequency would prove as strong a predictor of word knowledge as family frequency.

All three of the predictions were confirmed. All frequency by knowledge correlations were quite high, arithmetic grouping gave the highest correlations, and simple item frequency proved to be the strongest predictor of root word knowledge.

Word Frequency as a Predictor of English Root Word Knowledge

The problem of how to choose vocabulary items for instruction and/or testing purposes is an important issue in education. It is usually impractical to choose a large sample of words and then test a sample of the population of interest for knowledge of those words every time a test or lesson plan is made up. Therefore some measure that correlates reliably with word knowledge is desirable.

Graves, Ryder, Slater, & Calfee (1983) write that "... the most common method of selecting words for inclusion or exclusion in research studies, tests, or texts is to use word frequency as an index of word difficulty" (p. 2). Yet in many studies of the relationship between word frequency and word knowledge, a low correlation has been found between the frequency of a word in school English, and the number of students who show knowledge of that word on vocabulary tests. Graves, et al. (1983) report correlations ranging from .05 to .58 in such studies.

Noting the weak correlation between word frequency and word knowledge, Nagy and Anderson (1982) suggested that "family frequency" would prove to be a more reliable indicator. Family frequency is defined as total frequency of occurrence of the various forms of a word. It is the sum frequency of the members of a word family consisting of the root form and all of the inflected forms of a given word. For example, the word "called" is a member of the word family which also includes "call" and "caller," "calling," and "calls." The family frequency of "called" would then be the sum of the frequencies of all of these

words.

Graves, Ryder, Slater, & Calfee (1983) found a strong correlation between family frequency and school children's word knowledge, provided that the words are grouped, not taken one by one. With arithmetic grouping, they found a correlation of $r = .94$, and with a logarithmic grouping the correlation went up to $r = .97$. Thus, Graves et al. found that the family frequency of words, when the words are grouped by either method, is a good predictor of word knowledge.

The arithmetic grouping in the above study was done as follows. Once the frequencies of the test words have been determined, the items are ranked from least to most frequent. The items are then separated into blocks of 10 words each, and the mean percentage of subjects correctly identifying the words in each block is calculated. Each block is then assigned a number from one to k, and the mean percentages are correlated with the block numbers.

To illustrate logarithmic grouping, the first frequency block may contain the n most frequent words, the second the next 2n most frequent, the third the next 4n, and so on. The last block contains the remaining, least frequent words.

Using both forms of grouping, and looking at both individual item frequency and family frequency, Graves et al. (1983a) obtained the results shown in Table 1.

Insert Table 1 about here

Graves et al. conclude that given that groups of words

rather than individual words are considered, the frequency of the words can be an excellent predictor of children's knowledge of the words, and that both "family frequency and individual form frequency are equally good predictors" (1983b, p. 17).

In an earlier study, Graves (1980) brought up a point of importance to the present study. He wrote, "It seems reasonable that the likelihood that a derived form of a more common base would be known is better predicted by the frequency of the base or some weighted index derived from the frequency of the base and the frequency of the derived forms, than from the frequency of the derived forms alone" (p. 124). It seems likely, then, that the relationship between word frequency and word knowledge will be stronger for morphologically basic words with no affixes. The present study was therefore undertaken to determine if measures of the frequency of English root words can serve as reliable predictors of knowledge of those words.

The following hypotheses were tested.

(1) For morphologically basic (unaffixed) root words, there will be a strong relationship between word frequency and word knowledge.

(2) The advantage of logarithmic grouping over arithmetic grouping will be less apparent, or non-existent, for morphologically basic words.

(3) For grouped basic words, simple item frequency will prove as strong a predictor of word knowledge as family frequency.

Method

As part of a larger study of morphological knowledge (White,

Speidel, & Power, in preparation), two, 49 item vocabulary tests were developed. To select words for the test, a stratified random sample of prefixed words was drawn from The Word Frequency Book (Carroll, Davies, & Richman, 1971). The selected words were then stripped of all affixes down to the root form. Only English roots (defined as complete English words with meanings transparently related to the derivative) were included¹. The 98 items were then randomly assigned to one of two forms of the test.

Subjects

The tests were given to 46 third graders and 45 fourth graders from a private school serving predominantly middle class Hawaiian children in Honolulu, Hawaii, and 47 fourth graders from a public school in a lower SES, ethnically diverse community on the Island of Hawaii. There was a total of 138 subjects.

Procedure

The test was administered by a Caucasian male. The children took the test during their Language Arts period, and in their usual classrooms. Testing took approximately 40 minutes per class.

The tests were scored and the percentage of the subjects who correctly identified the meaning of each root word was determined. Four measures of root-word frequency were calculated using The Word Frequency Book (Carroll et al., 1971).

f: Simple frequency of each word. Item frequency.

u: Item frequency per million words of running text.

Ff: Family frequency.

Fu: Sum of the u values for each family.

Family membership was determined by use of the following criteria developed by Graves, et al. (1983, p.9):

- Include all forms which are likely to contribute to students' learning of the meaning of the target word tested.
- A base word (root) and all of its suffixed forms constitute a family.
- Only forms of a compound word that contain both words of the compound are considered to be members of a family.
- Prefixed words are considered to constitute a different family than their non-prefixed relatives.

A scatterplot of the data revealed that two of the items, "word" and "man", were extreme outliers. Thus the frequencies of these forms may be misleading. They were therefore removed from the data set, leaving 95 items, the original 98 minus the non-root² "partially," and the misleading items "word" and "man."

Results

Ungrouped Data

For each of the four measures of frequency, a Pearson r was calculated between word frequency and the percentage of children correctly responding on the multiple choice vocabulary test. The items were not grouped, but compared one to one, percentage to frequency. The results are shown in Table 2.

 Insert Table 2 about here

Clearly the ungrouped data do not show strong correlation with word knowledge.

Grouped Data

Next, the items were grouped arithmetically, as discussed above (p. 4). The correlation coefficients for each of the four measures of frequency are shown in Table 3.

Insert Table 3 about here

Finally, when the items were grouped logarithmically as discussed above (p. 4), the results shown in Table 4 were obtained.

Insert Table 4 about here

Discussion

The best predictor of knowledge of root words in this study was simple item frequency (f), using an arithmetic grouping. Arithmetic grouping of frequency adjusted for dispersion (u) and the two measures of family frequency (Ff and Fu) were all found to be more strongly correlated with word knowledge than any of the logarithmic groupings. A ranking of the measures of frequency from those showing the strongest to those showing the least correlation is given in Table 5.

Insert Table 5 about here

The hypothesis that the data would show a strong relationship between word frequency and word knowledge for root words was clearly supported, as was the hypothesis that simple item frequency would not give markedly different results from family frequency. There was not much difference between the

results obtained through arithmetic and logarithmic grouping schemes, however logarithmic grouping of all frequency measures was lower than arithmetic, supporting the third hypothesis.

As previous research has shown, the frequency of an individual word was found to be a relatively weak predictor of children's knowledge of that word. However, when the words are dealt with in groups, with a few as ten words of similar frequencies in each group, the correlation between frequency and knowledge rises to as much as $r = .98$, depending on the measure of frequency used. These findings may have particular implications for morphological instruction, as it is useful to know which root words are likely to be known by the students for the purposes of illustrating word formation processes and affixation.

References

- Carroll, J. B., Davies, P., & Richman, B. (1971). The American heritage word frequency book. Boston: Houghton Mifflin.
- Graves, M. F., Boettcher, J. A., & Peacock, J. L. (1980). Word frequency as a predictor of students' reading vocabularies. Journal of Reading Behavior, 12(2), 117-127.
- Graves, M. F., Ryder, R. J., Slater, W. H., & Calfee, R. C. (1983, April). The relationship between word frequency and reading vocabularies using four metrics of frequency. Paper presented at convention of the American Educational Research Association, Montreal, Canada.
- Nagy, W. E., & Anderson, R. C. (1982). The number of words in printed school English (Tech. Rep. No. 253). Champaign: University of Illinois, Center for the Study of Reading.
- Nagy, W. E., & Anderson, R. C. (1984). How many words in printed school English? Reading Research Quarterly, 19, 304-330.
- White, T. G., Spidel, G. E., & Power, M. A. (In preparation). Morphological analysis: A useful strategy for elementary students? Honolulu: Kamehameha Schools, Center for Development of Elementary School.

Footnotes

¹ This criterion led to dropping one item from inclusion in the data analysis due to an affixed form having mistakenly been used instead of the root ("partial" instead of "part").

² The randomly chosen derived forms were "unmanned," and "reword." After being stripped of the prefixes, the two root words were included in the test as verbs, according to the selection criteria of knowledge of the root contributing to learning of the derivative. This was valid from the viewpoint of the morphology study, but as The Word Frequency Book does not discriminate between syntactic forms, the more common noun forms of the words were felt to be responsible for the high frequency counts. The less common verb forms were not well known (word = 12%, man = 18%).

It should be noted that these two outliers, "word" and "man" could have had a great deal of influence on the correlations. When they are included in the data, the correlations go to near zero for all measures of frequency. In future research, care must be taken to assure the part of speech tested is the one responsible for the frequency count for that word.

Table 1

Correlations Between Metrics of Frequency and Word Knowledge

(Adapted from Graves [1983b; pp13-14].)

<u>Frequency Measure</u>	<u>Grouping Scheme</u>	
	<u>Arithmetic</u>	<u>Logarithmic</u>
Item frequency	$r=.68$	$r=.99$
Family frequency	$r=.92$	$r=.97$

Table 2

Correlations Without Grouping

<u>Frequency Measure</u>	<u>Correlation With</u> <u>% Correct</u>
f	.35
u	.33
Ff	.38
Fu	.35

Table 3

Correlations with Arithmetic Grouping

<u>Frequency Measure</u>	<u>Correlation With</u> <u>% Correct</u>
f	.98
u	.92
Ff	.96
Fu	.94

Table 4

Correlations With Logarithmic Grouping

<u>Frequency Measure</u>	<u>Correlation With</u> <u>% Correct</u>
f	.89
u	.90
Ff	.89
Fu	.83

Table 5

Rank Order of Correlations

<u>Grouping Scheme</u>	Correlation With <u>% Correct</u>
f - arithmetic	.98
Ff - arithmetic	.96
Fu - arithmetic	.94
u - arithmetic	.92
u - logarithmic	.90
f/Ff - logarithmic	.89
Fu - logarithmic	.83